International Journal of Innovative Research in Engineering & Management (IJIREM)
ISSN: 2350-0557, Volume-9, Issue-2, April 2022
https://doi.org/10.55524/ijirem.2022.9.2.11
Article ID IJIR-1194, Pages 84-90
www.ijirem.org

Analysis of Driver Drowsiness Using Convolution Neural Network Algorithms

Gazee Afzal Wani[1], Ravinder PaL Singh[2], and Dr. Monika Mehra[3]

[1]M. Tech Scholar, Department of Electronics and Communication Engineering, RIMT University, Punjab, India
[2]Technical Head, Department of Research, Innovation & Incubation, RIMT University, Punjab, India
[3]Head of Department, Department of Electronics and Communication Engineering, RIMT University, Punjab, India

Correspondence should be addressed to Gazee Afzal Wani; gazeewani789@gmail.com

ABSTRACT- Our security is the need while voyaging or driving. One slip-up of the driver can prompt serious actual wounds, passing and critical monetary misfortunes. These days there are numerous frameworks accessible in market like route frameworks, different sensors and so forth to make driver's work simple. There are different reasons particularly human blunders which gives ascends to the street mishaps. Reports say that there is an enormous addition in the street mishaps in our country since most recent couple of years. The principal reason happening from the interstate mishaps is the laziness and tiredness of driver while driving. It is a vital stage to accompany a proficient procedure to recognize sleepiness when driver feels tired. This could save enormous number of mishaps to happen. In this framework, we proposed to decrease the quantity of mishaps brought about by driver weariness and hence further develop street wellbeing. We find, track, and break down both the driver face and eyes to quantify PERCLOS (level of eye conclusion).

KEYWORDS- PERCLOS, AI, Driver weakness, driver drowsiness, CNN Algorithm

I. INTRODUCTION

If you have driven previously, you've been languid in the driver's seat eventually. It's not something we like to concede but rather it's a significant issue with genuine outcomes that should be tended to. 1 out of 4 vehicle mishaps are brought about by tired driving and 1 out of 25 grown-up drivers report that they have nodded off at the worst possible time in the beyond 30 days. The most alarming part is that sleepy driving isn't simply nodding off while driving. Languid driving can be pretty much as little as a concise condition of obviousness when the driver isn't giving full consideration to the street. Tired driving outcomes in more than 71,000 wounds, 1,500 passing, and $12.5 billion in money related misfortunes each year. Because of the pertinence of this issue, we accept it is vital to foster an answer for laziness identification, particularly in the beginning phases to forestall mishaps.[1][2]

Also, we accept that sleepiness can contrarily affect individuals in working and study hall conditions too. Even though lack of sleep and school go connected at the hip, laziness in the work environment particularly while working with large equipment might bring about genuine wounds like those that happen while driving languidly.

Our answer for this issue is to assemble an identification framework that distinguishes key ascribes of languor and triggers a ready when somebody is lazy before it is past the point of no return.

II. DATA SOURCE AND PREPROCESSING

For our preparation and test information, we utilized the Genuine Tiredness Dataset made by an examination group from the College of Texas at Arlington explicitly for identifying multi-stage laziness. The ultimate objective is to recognize outrageous and apparent instances of languor as well as permit our framework to identify gentler signs of sluggishness too. The dataset comprises of around 30 hours of recordings of 60 special members. From the dataset, we had the option to extricate facial milestones from 44 recordings of 22 members. This permitted us to get an adequate measure of information for both the ready and tired state.

For each video, we used OpenCV to extract 1 frame per second starting at the 3-minute mark until the end of the video.

Each video was approximately 10 minutes long, so we extracted around 240 frames per video, resulting in 10560 frames for the entire dataset.

```
import cv2
data = []
labels = []
for j in [60]:
  for i in [10]:
    vidcap          =          cv2.VideoCapture(‘drive/My
Drive/Fold5_part2/’ +     str(j) +’/’ + str(i) + ‘.mp4’)
sec = 0
    frameRate = 1
    success, image = getFrame(sec)
    count = 0
    while success and count < 240:
      landmarks = extract_face_landmarks(image)
      if sum(sum(landmarks)) != 0:
        count += 1
        data.append(landmarks)
        labels.append([i])
        sec = sec + frameRate
        sec = round(sec, 2)
        success, image = getFrame(sec)
        print(count)
    else:
        sec = sec + frameRate
        sec = round(sec, 2)
        success, image = getFrame(sec)
        print(“not detected”)
```

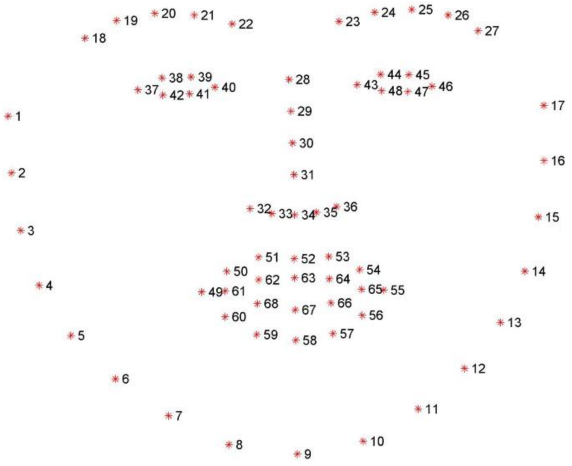


Figure 1: Facial Landmarks from OpenCV

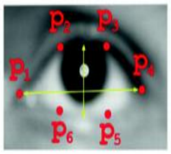
There were 68 all out milestones for every casing except we chose to save the tourist spots for the eyes and mouth as shown in fig 1, just (Focuses 37-68). These were the significant information focuses we used to separate the elements for our model.[3]

III. FEATURE EXTRACTION

As momentarily implied before, in view of the facial milestones that we extricated from the casings of the recordings, we wandered into creating appropriate highlights for our arrangement model. While we conjectured and tried a few highlights, the four center elements that we closed on for our last models were eye viewpoint proportion, mouth angle proportion, student circularity, lastly, mouth perspective proportion over eye viewpoint proportion.

A. Eye Aspect Proportion (EAR)

EAR, as the name recommends, is the proportion of the length of the eyes to the width of the eyes. The length of the eyes is determined by averaging north of two unmistakable vertical lines across the eyes as shown in the figure 2 underneath.



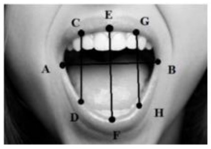
$$\text{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

Figure 2: Eye Aspect Ratio (EAR)

Our theory was that when an individual is lazy, their eyes are probably going to get more modest, and they are probably going to flicker more. In view of this theory, we anticipated that our model should foresee the class as tired assuming the eye perspective proportion for a person over progressive edges began to decrease for example their eyes began to be more shut, or they were squinting quicker.

B. Mouth Aspect Ratio (MAR)

Computationally like the EAR, the Blemish, as you would anticipate, measures the proportion of the length of the mouth to the width of the mouth. Our speculation was that as an individual becomes tired, they are probably going to yawn and let completely go over their mouth, making their Blemish to be higher than expected in this state as in fig 3.



$$\text{MAR} = \frac{|\text{EF}|}{|\text{AB}|}$$

Figure 3: Mouth aspect ratio

C. Pupil Circulatory (PUC)

As in fig 4, PUC is an action reciprocal to EAR, yet it puts a more prominent accentuation on the student rather than the whole eye.

$$Circularity = \frac{4 * \pi * Area}{perimeter^2} \qquad Area = \left(\frac{\text{Distance(p2, p5)}}{2}\right)^2 * \pi$$

$$\begin{aligned} Perimeter = {} & Distance(p1,p2) + Distance(p2,p3) + Distance(p3,p4) + \\ & Distance(p4,p5) + Distance(p5,p6) + Distance(p6,p1) \end{aligned}$$

Figure 4: Pupil Circularity

For instance, somebody who has their eyes half-open or nearly shut will have a much lower student circularity esteem versus somebody who has their eyes completely open because of the squared term in the denominator. Like the EAR, the assumption was that when an individual is sluggish, their understudy circularity is probably going to decay.

D. Mouth aspect ratio over Eye aspect ratio (MOE)

At long last, we chose to add MOE as another component. MOE is just the proportion of the Blemish to the EAR.

$$MOE = \frac{MAR}{EAR}$$

Mouth aspect ratio over Eye aspect ratio (MOE)
The advantage of utilizing this component is that EAR and Blemish are relied upon to move in inverse bearings if the condition of the singular changes. Instead of both EAR and Blemish, MOE as an action will be more receptive to these progressions as it will catch the unpretentious changes in both EAR and Blemish and will misrepresent the progressions as the denominator and numerator move in inverse bearings. Since the MOE accepts Blemish as the numerator and EAR as the denominator, our hypothesis was that as the individual gets sluggish, the MOE will increment. While this multitude of highlights appeared to be legit, when tried with our arrangement models, they yielded helpless outcomes in the scope of 55% to 60% exactness which is just a minor improvement over the pattern precision of half for a paired adjusted order issue. Regardless, this failure drove us to our most significant disclosure: the highlights were right, we simply weren't checking out them accurately.

IV. FEATURE NORMALIZATION

At the point when we were trying our models with the four center elements talked about above, we saw a disturbing example. At whatever point we arbitrarily split the edges in our preparation and test, our model would yield results with precision as high 70%, notwithstanding, at whatever point we split the casings by people (for example a person that is in the test set won't be in the preparation set), our model exhibition would be poor as insinuated before.
This drove us to the acknowledgment that our model was battling with new faces and the essential justification for this battle was the way that every individual has different center elements in their default ready state. That is, individual A may normally have a lot more modest eyes than individual B. On the off chance that a model is prepared on individual B, the model, when tried on individual A, will forever anticipate the state as languid because it will distinguish a fall in EAR and PUC and an ascent in MOE even though individual A was ready. In view of this disclosure, we conjectured that normalizing the elements for every individual is probably going to yield better outcomes and as it ended up, we were right.[4]
To standardize the elements of every person, we took the initial three casings for every individual's ready video and involved them as the pattern for standardization. The mean and standard deviation of each element for these three edges were determined and used to standardize each element separately for every member. Numerically, this is what the standardization condition resembled:

$$Normalised\ Feature_{n,m} = \frac{Feature_{n,m} - \mu_{n,m}}{\sigma_{n,m}}$$

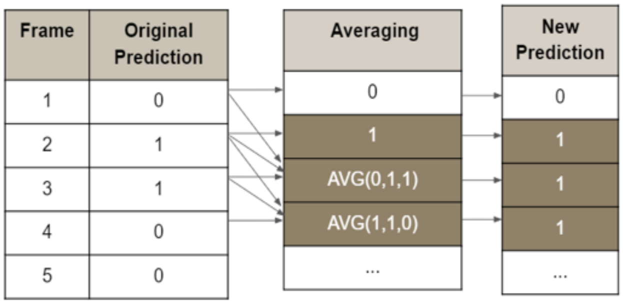
where:
n is the feature
m is the person
$\mu_{n,m}$ and $\sigma_{n,m}$ are taken from the first 3 frames of the "Alert" state

A. Normalization Method

Since we had standardized every one of the four center elements, our list of capabilities had eight highlights, each center component supplemented by its standardized variant. We tried every one of the eight elements in our models and our outcomes improved altogether.

B. Basic Classification Methods and Results

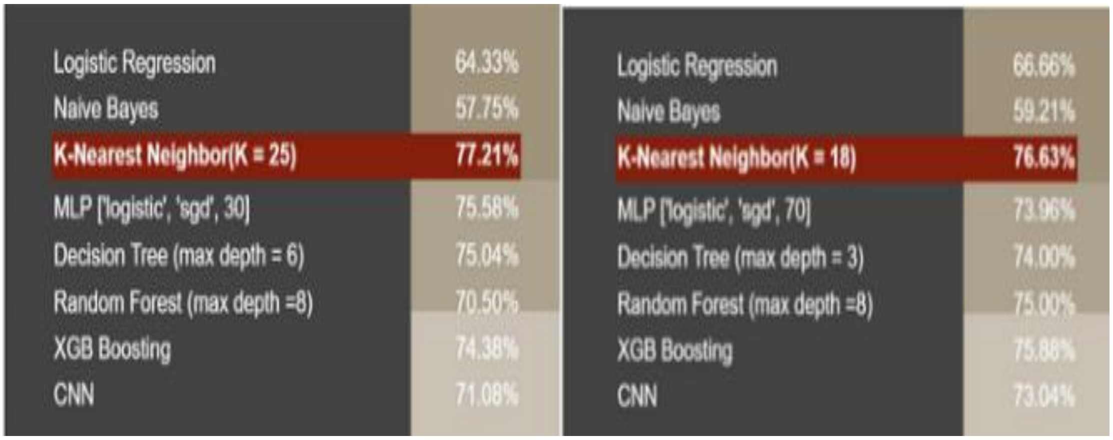
After we separated and standardized our highlights, we needed to attempt a progression of displaying procedures, beginning with the most essential characterization models like strategic relapse and Credulous Bayes, continuing to more complicated models containing neural organizations and other profound learning draws near. It's critical to take note of the presentation interpretability tradeoff here. Although we focus on top-performing models, interpretability is additionally critical to us if we somehow happened to popularize this arrangement and present its business suggestions to partners who are curious about the AI language. To prepare and test our models, we split our dataset into information from 17 recordings and information from 5 recordings individually. Thus, our preparation dataset contains 8160 lines, and our test dataset contains 2400 columns.
How would we acquaint arrangement with fundamental characterization techniques?
One test we looked during this task was that we were attempting to anticipate the name for each edge in the succession. While complex models like LSTM and RNN can represent successive information, fundamental grouping models can't.
The way we managed this issue was to average the first expectation results with the forecast outcomes from the past two edges. Since our dataset was partitioned into preparing and test considering the singular members and the information focuses are all in the request for time succession, averaging seems OK for this situation and permitted us to convey more exact forecasts as depicted in fig 5.



Frame	Original Prediction	Averaging	New Prediction
1	0	0	0
2	1	1	1
3	1	AVG(0,1,1)	1
4	0	AVG(1,1,0)	1
5	0	...	...

Figure 5: Acquainting Arrangement with Fundamental Characterization Models

From the different grouping techniques we attempted, K-Closest Neighbor (kNN, k = 25) had the most noteworthy out-of-test exactness of 77.21%. Gullible Bayes played out the most awful at 57.75% and we presumed that this was on the grounds that the model makes some harder memories managing mathematical information. Although kNN yielded the most elevated exactness, the bogus negative rate was very high at 0.42 which intends that there is a 42% likelihood that somebody who is really sleepy would be distinguished as ready by our framework. To diminish the bogus negative rate, we brought the edge from 0.5 down to 0.4 which permitted our model to anticipate a larger number of cases lazy than alert. Albeit the exact nesses for a portion of different models expanded, kNN actually detailed the most noteworthy precision at 76.63% (k = 18) regardless of a decrease in its own exactness as in fig 6.



Logistic Regression	64.33%
Naive Bayes	57.75%
K-Nearest Neighbor(K = 25)	77.21%
MLP ['logistic', 'sgd', 30]	75.58%
Decision Tree (max depth = 6)	75.04%
Random Forest (max depth =8)	70.50%
XGB Boosting	74.38%
CNN	71.08%

Logistic Regression	66.66%
Naive Bayes	59.21%
K-Nearest Neighbor(K = 18)	76.63%
MLP ['logistic', 'sgd', 70]	73.96%
Decision Tree (max depth = 3)	74.00%
Random Forest (max depth =8)	75.00%
XGB Boosting	75.88%
CNN	73.04%

Figure 6: Left: Original Results

Right: Results after lowering threshold from 0.5 -> 0.4

C. Feature Importance

We needed to get a feeling of component significance so we imagined the outcomes from our Arbitrary Woods model shown in figure 7.

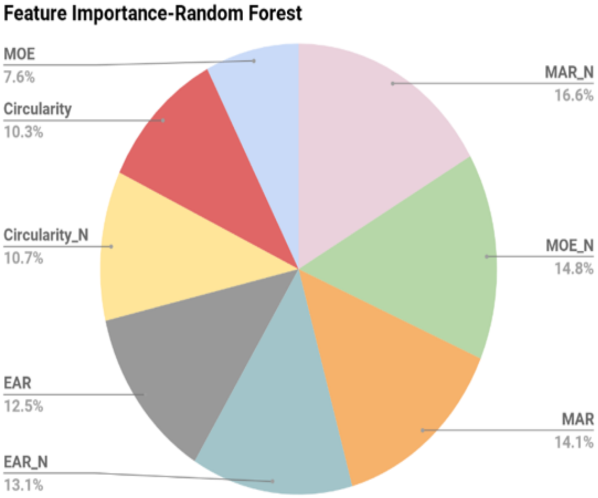


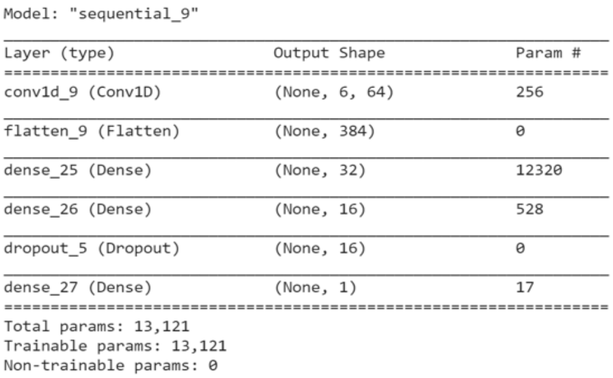
Figure 7: Feature Importance

Mouth Perspective Proportion after standardization ended up being the main component out of our 8 highlights. This appears to be legit on the grounds that when we are sluggish, we will quite often yawn even more regularly. Normalizing our highlights misrepresented this impact and made it a superior sign of tiredness in various members.[4]

V. CONVOLUTIONAL NEURAL ORGANIZATION (CNN)

Convolutional Neural Organizations (CNN) are ordinarily used to dissect picture information and guide pictures to yield factors. Be that as it may, we chose to assemble a 1-D CNN and send in mathematical elements as successive info information to attempt to comprehend the spatial connection between each element for the two states. Our CNN model has 5 layers including 1 convolutional layer, 1 straighten later, 2 completely associated thick layers, and 1 dropout layer before the result layer. The straighten layer smooths the result from the convolutional layer and makes it direct prior to passing it into the primary thick layer. The dropout layer haphazardly drops 20% of the result hubs from the second thick layer to keep our model from overfitting to the preparation information. The last thick layer has a solitary result hub that yields 0 for ready and 1 for languid as in table 1.

Table 1: CNN Model Design



```
Model: "sequential_9"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_9 (Conv1D)            (None, 6, 64)             256
_________________________________________________________________
flatten_9 (Flatten)          (None, 384)               0
_________________________________________________________________
dense_25 (Dense)             (None, 32)                12320
_________________________________________________________________
dense_26 (Dense)             (None, 16)                528
_________________________________________________________________
dropout_5 (Dropout)          (None, 16)                0
_________________________________________________________________
dense_27 (Dense)             (None, 1)                 17
=================================================================
Total params: 13,121
Trainable params: 13,121
Non-trainable params: 0
```

Activation Function	Relu/Sigmoid
Optimizer	Adam
Loss Function	Binary Crossentropy
Number of Epochs	100
Learning Rate	0.00001

VI. LONG SHORT-TERM MEMORY (LSTM) NETWORKS

One more strategy to manage successive information is utilizing a LSTM model. LSTM networks are a unique sort of Repetitive Neural Organizations (RNN), equipped for learning long haul conditions in the information. Repetitive Neural Organizations are input neural organizations that have inside memory that permits data to persevere.

How might RNNs have an inward memory space while handling new information?

The response is that when settling on a choice, RNNs consider the current contribution as well as the result that it has gained from the past sources of info. This is likewise the primary distinction among RNNs and other neural organizations. In other neural organizations, the data sources are autonomous of one another. In RNNs, the data sources relate to one another. The equation is as underneath:

$$h_t = f(h_{t-1}, x_t)$$

RNN formula

We decided to utilize a LSTM network since it permits us to concentrate on long successions without stressing over the inclination evaporating issues looked by conventional RNNs. Inside the LSTM organization, there are three entryways for each time step: Neglect Door, Information door, and Result Entryway as shown in fig 8.

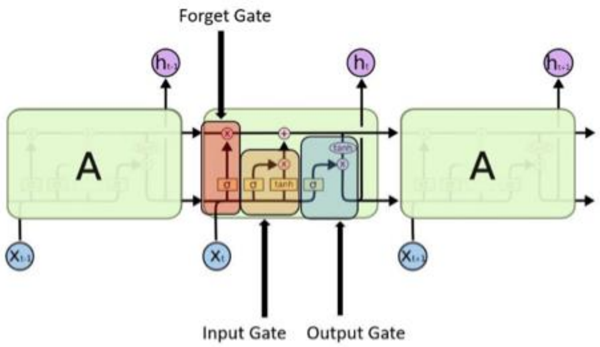


Figure 8: LSTM Network Visualized

Disregard Door: as its name proposes, the entryway attempts to "neglect" part of the memory from the past result.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

Forget Gate Formula

Input Gate: The entryway concludes what ought to be kept from the contribution to request to alter the memory.

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

input Gate Equation

Output Gate: The entryway concludes what the result is by joining the information and memory.

$$o_t = \sigma\left(W_o \ [h_{t-1}, x_t] \ + \ b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

Output Gate Equation

In the first place, we changed over our recordings into bunches of information. Then, at that point, each bunch was sent through a completely associated layer with 1024 secret units utilizing the sigmoid enactment work. The following layer is our LSTM layer with 512 secret units followed by 3 additional FC layers until the last result layer as shown in fig 9.

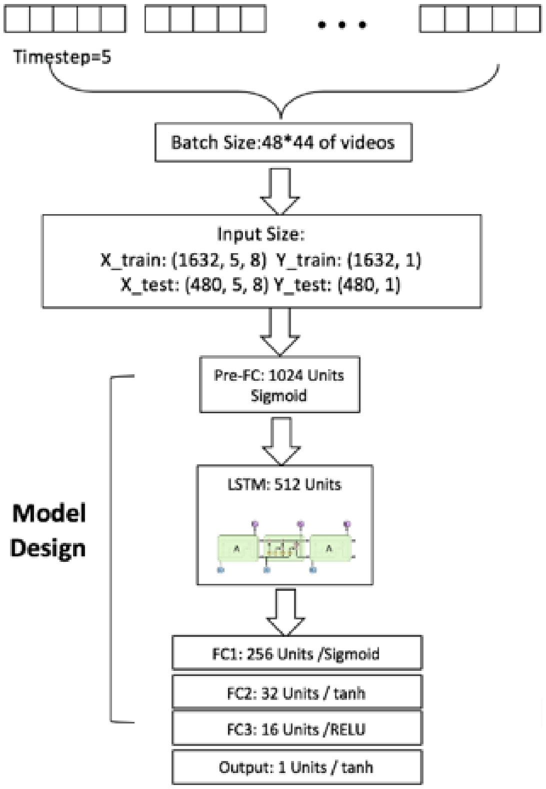
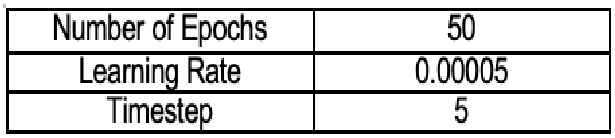


Figure 9: LSTM Network Design

Table 2: LSTM Parameters



Number of Epochs	50
Learning Rate	0.00005
Timestep	5

After hyperparameter tuning, our streamlined LSTM model accomplished a general precision of 77.08% with a much lower bogus negative pace of 0.3 contrasted with the bogus negative pace of our kNN model (0.42) as depicted by table 2.[5]

VII. TRANSFER LEARNING

Transfer learning centers around utilizing the information acquired while tackling one issue and applying it to take care of an alternate however related issue.As fig 10 shows It is a valuable arrangement of strategies particularly for situations when we have restricted opportunity to prepare the model or restricted information to completely prepare a neural organization. Since the information we were working with had not many interesting examples, we accepted this issue would be a decent possibility for utilizing move learning. The model we chose to utilize is VGG16 with the ImageNet dataset.

VGG16 is a convolutional neural organization model which was proposed by K. Simonyan and A. Zisserman from the College of Oxford in their paper "Exceptionally Profound Convolutional Organizations for Enormous Scope Picture Acknowledgment". The model figured out how to accomplish 92.7% top-5 test exactness in ImageNet, which is a dataset of more than 14 million pictures having a place with 1000 classes.

ImageNet is a dataset with north of 15 million named high-goal pictures having a place with around 22,000 unique classifications. The pictures were gathered from the web and marked by human labelers utilizing Amazon's publicly supporting device, Mechanical Turk. Beginning around 2010, as a component of the Pascal Visual Article Challenge, a contest called the ImageNet Huge Scope Visual Acknowledgment Challenge (ILSVRC) is held every year. ILSVRC utilizes a more modest arrangement of ImageNet with around 1000 pictures in every one of 1000 classifications. There are roughly 1.2 million preparation pictures, 50,000 approval pictures, and 150,000 testing pictures. ImageNet comprises of pictures with various goals. Consequently, the goal of pictures should be changed to a proper worth of 256×256. The picture is rescaled and edited out and the focal 256×256 fix frames the subsequent picture. The contribution to cov1 layer is a 224 x 224 RGB picture. The picture is gone through a heap of convolutional layers,where the channels are utilized with a tiny open field: 3×3. In one of the designs, the model additionally uses 1×1 convolution channels, which should be visible as a direct change of the info channels followed by non-straight changes. The convolution step is fixed to 1 pixel; the spatial cushioning of convolutional layer input is to such an extent that the spatial goal is saved after convolution, for example the cushioning is 1-pixel for 3×3 convolutional layers. Spatial pooling is completed by five max-pooling layers, which follow a portion of the convolutional layers. Not all the conv. layers are trailed by max-pooling. Max-pooling is performed over a 2×2 pixel window, with a step of 2.

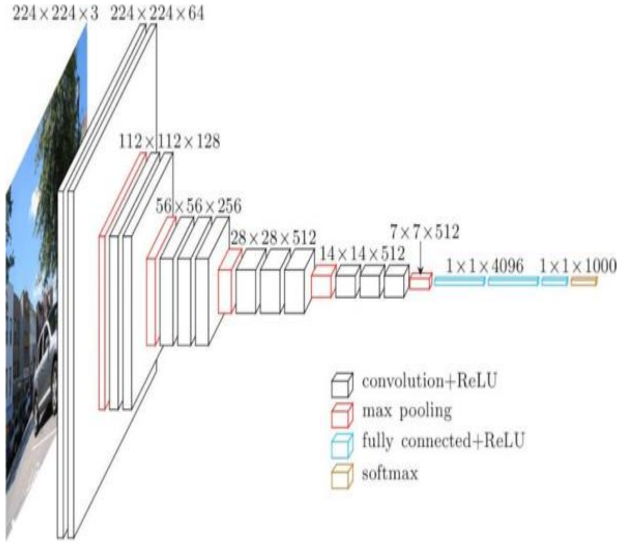


Figure 10: VGG16 Network Architecture

Three Completely Associated (FC) layers follow a pile of convolutional layers: the initial two have 4096 channels each, the third performs 1000-way ILSVRC order and accordingly contains 1000 channels. The last layer is a delicate max layer. The setup of the completely associated layers is something similar in all organizations.

All secret layers are furnished with the amendment (ReLU) non-linearity. It is additionally noticed that notwithstanding one none of the organizations contain Neighborhood Reaction Standardization (LRN), on the grounds that such standardization doesn't work on the presentation of the model, however prompts expanded calculation time.

We split the preparation recordings into 34,000 pictures which were screen captures taken each 10 casings. We took care of these pictures to the VGG16 model. We accepted that the quantity of pictures was adequate to prepare the pre-prepared model. We got the accompanying exactness scores in the wake of preparing the model for 50 ages. Our outcomes are displayed in fig 11.

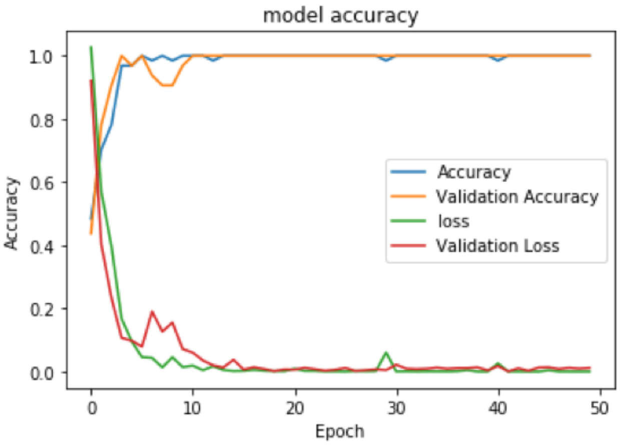


Figure 11: VGG16 Results

Obviously, the model was overfitting. A potential clarification for this is that pictures that we went through the model was of 22 respondents sitting for all intents and purposes still before a camera with undisturbed foundations. So notwithstanding taking countless edges (34,000) into our model, the model was basically attempting to gain from 22 arrangements of for all intents and purposes indistinguishable pictures. Consequently, the model didn't actually have sufficient preparation information from a genuine perspective.[6]

VIII. CONCLUSION AND FUTURE SCOPE

We advanced many things all through this task. In the first place, easier models can be similarly as proficient at doing jobs as more perplexing models. For our situation, the K-Closest Neighbor model gave an exactness like the LSTM model. In any case, since we would rather not misclassify individuals who are sleepy as ready, eventually it is smarter to utilize the more perplexing model with a lower bogus negative rate than a less difficult model that might be less expensive to convey. Second, standardization was essential to our presentation. We perceived that everyone has an alternate gauge for eye and mouth angle proportions and normalizing for every member was important. Outside of runtime for our models, information pre-handling and component extraction/standardization took up a greater part within recent memory. It will be intriguing to refresh our venture and investigate how we can diminish the bogus negative rate for kNN and other less difficult models.

Moving, there are a couple of things we can do to additionally work on our outcomes and calibrate the models. To begin with, we want to fuse distance between the facial milestones to represent any development by the subject in the video. Sensibly the members won't be static on the screen, and we accept unexpected developments by the member might flag tiredness or awakening from miniature rest. Second, we need to refresh boundaries with our more mind-boggling models (NNs, groups, and so on) to accomplish better outcomes. Third lastly, we might want to gather our own preparation information from a bigger example of members (more data!!!) while including new unmistakable signs of tiredness like unexpected head development, hand development, or in any event, following eye developments.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Marco Javier Flores • José María Armingol • Arturo de la Escalera.: Real-Time Warning System for Driver Drowsiness Using Visual Information. In: Springer Science + Business Media B.V. 2009

[2] Luis M. Bergasa, Jesús Nuevo, Miguel A. Sotelo, RafaelBarea, and María Elena Lopez.:Real-Time System forMonitoring Driver Vigilance. In: ieee transactions on intelligent transportation systems, vol. 7, no. 1, march 2006

[3] Mohamad-Hoseyn Sigari, Mahmood Fathy, and Mohsen Soryani.: A Driver Face Monitoring System for Fatigue and Distraction Detection. In: Hindawi Publishing Corporation International Journal of Vehicular Technology, Volume 2013, Article ID 263983, 11 pages

[4] Jay D. Fuletra, Bulari Bosamia: A Survey On Driver's Drowsiness Detection Techniques presented at IJRITCC in November 2013.

[5] Ming-ai Li, Cheng Zhang, Jin-Fu Yang. :An EEG-based Method for Detecting Drowsy Driving State. In: Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on , 5, pp. 2164- 2167, 10-12 Aug. 2010.

[6] Er. Manoram Vats and Er. Anil Garg.: Detection And Security System For Drowsy Driver By Using Artificial Neural Network Technique. In: International Journal of Applied Science and Advance Technology January-June 2012, Vol. 1, No. 1, pp. 39-43

[7] Examining individual differences. J. Sleep Res. 2006, 15, 47–53.